

# Frequency Offset Estimation and Training Sequence Design for MIMO OFDM

Yanxiang Jiang, *Member, IEEE*, Hlaing Minn, *Senior Member, IEEE*, Xiqi Gao, *Senior Member, IEEE*, Xiaohu You, *Member, IEEE*, and Yinghui Li, *Student Member, IEEE*

**Abstract**—This paper addresses carrier frequency offset (CFO) estimation and training sequence design for multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) systems over frequency selective fading channels. By exploiting the orthogonality of the training sequences in the frequency domain, integer CFO (ICFO) is estimated. With the uniformly spaced non-zero pilots in the training sequences and the corresponding geometric mapping, fractional CFO (FCFO) is estimated through the roots of a real polynomial. Furthermore, the condition for the training sequences to guarantee estimation identifiability is developed. Through the analysis of the correlation property of the training sequences, two types of sub-optimal training sequences generated from the Chu sequence are constructed. Simulation results verify the good performance of the CFO estimator assisted by the proposed training sequences.

**Index Terms**—MIMO-OFDM, frequency selective fading channels, training sequences, frequency offset estimation.

## I. INTRODUCTION

ORTHOGONAL frequency division multiplexing (OFDM) transmission is receiving increasing attention in recent years due to its robustness to frequency-selective fading and its subcarrier-wise adaptability. On the other hand, multiple-input multiple-output (MIMO) systems attract considerable interest due to the higher capacity and spectral efficiency that they can provide in comparison with single-input single-output (SISO) systems. Accordingly, MIMO-OFDM has emerged as a strong candidate for beyond third generation (B3G) mobile wide-band communications [1].

It is well known that SISO-OFDM is highly sensitive to carrier frequency offset (CFO), and accurate estimation and

compensation of CFO is very important [2]. A number of approaches have dealt with CFO estimation in a SISO-OFDM setup [3], [4], [2], [5], [6], [7]. According to whether the CFO estimators use training sequences or not, they can be classified as blind ones [3] [4] and training-based ones [2], [5], [6], [7]. Similar to SISO-OFDM, MIMO-OFDM is also very sensitive to CFO. Moreover, for MIMO-OFDM, there exists multi-antenna interference (MAI) between the received signals from different transmit antennas. The MAI makes CFO estimation more difficult, and a careful training sequence design is required for training-based CFO estimation. However, unlike SISO-OFDM, only a few works on CFO estimation for MIMO-OFDM have appeared in the literature. In [8], a blind kurtosis-based CFO estimator for MIMO-OFDM was developed. For training-based CFO estimators, the overviews concerning the necessary changes to the training sequences and the corresponding CFO estimators when extending SISO-OFDM to MIMO-OFDM were provided in [9], [10]. However, with the provided training sequences in [9], satisfactory CFO estimation performance cannot be achieved. With the training sequences in [10], the training period grows linearly with the number of transmit antennas, which results in an increased overhead. In [11], a white sequence based maximum likelihood (ML) CFO estimator was addressed for MIMO, while a hopping pilot based CFO estimator was proposed for MIMO-OFDM in [12]. Numerical calculations of the CFO estimators in [11] [12] require a large point discrete Fourier transform (DFT) operation and a time consuming line search over a large set of frequency grids, which make the estimation computationally prohibitive. To reduce complexity, computationally efficient CFO estimation was introduced in [13] by exploiting proper approximations. However, the CFO estimator in [13] is only applied to flat-fading MIMO channels.

When training sequence design for CFO estimation is concerned, it has received relatively little attention. It was investigated for single antenna systems in [14], where a white sequence was found to minimize the worst-case asymptotic Cramer-Rao bound (CRB). Recently, an improved training sequence and structure design was developed in [15] by exploiting the CRB and received training signal statistics. In [16], training sequences were designed for CFO estimation in MIMO systems using a channel-independent CRB. In [17], the effect of CFO was incorporated into the mean-square error (MSE) optimal training sequence designs for MIMO-OFDM channel estimation in [18]. Note that optimal training sequence design for MIMO-OFDM CFO estimation

Manuscript received October 18, 2006; revised February 5, 2007; accepted February 10, 2007. The associate editor coordinating the review of this manuscript and approving it for publication was D. Huang. The work of Yanxiang Jiang, Xiqi Gao, and Xiaohu You was supported in part by the National Natural Science Foundation of China under Grants 60496311 and 60572072, the China High-Tech 863 Project under Grant 2003AA123310 and 2006AA01Z264, and the International Cooperation Project on Beyond 3G Mobile of China under Grant 2005DFA10360. The work of Hlaing Minn and Yinghui Li was supported in part by the Erik Jonsson School Research Excellence Initiative, the University of Texas at Dallas, USA. This paper was presented in part at the IEEE International Conference on Communications (ICC), Istanbul, Turkey, June 2006.

Y. Jiang, X. Gao, and X. You are with the National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China (e-mail: {yxjiang, xqgao, xhyu}@seu.edu.cn).

H. Minn and Y. Li are with the Department of Electrical Engineering, University of Texas at Dallas, TX 75083-0688, USA (e-mail: {hlaing.minn, yinghui.li}@utdallas.edu).

Digital Object Identifier 10.1109/TWC.2008.060846.

in frequency selective fading channels is still an open problem.

In this paper, we propose a new training-based CFO estimator for MIMO-OFDM over frequency selective fading channels. To avoid the large point DFT operation and the time consuming line search and also to guarantee the estimate precision, we propose to first estimate integer CFO (ICFO) and then fractional CFO (FCFO). Next, we relate FCFO estimation to direction of arrival (DOA) estimation in uniform linear arrays (ULA). Then, we propose to exploit a geometric mapping to transform the complex polynomial related to FCFO estimation into a real polynomial. Correspondingly, FCFO is estimated through the roots of the real polynomial. In the derivation of the CFO estimator, we also develop the identifiability condition concerning the training sequences. Furthermore, exploiting the well-studied results for DOA estimation, we transform the problem of training sequence design for CFO estimation into the study of the correlation property of the training sequences, and propose to construct them from the Chu sequence [19].

The rest of this paper is organized as follows. In Section II, we briefly describe the MIMO-OFDM system model. The proposed CFO estimator including ICFO and FCFO estimation is presented in Section III. The aspect concerning training sequence design is addressed in Section IV. Simulation results are shown in Section V. Final conclusions are drawn in Section VI.

*Notations:* Upper (lower) bold-face letters are used for matrices (column vectors). Superscripts  $*$ ,  $T$  and  $H$  denote conjugate, transpose and Hermitian transpose, respectively.  $(\cdot)_P$  denotes the residue of the number within the brackets modulo  $P$ .  $\Re(\cdot)$  and  $\Im(\cdot)$  denote the real and imaginary parts of the enclosed parameters, respectively.  $\lfloor \cdot \rfloor$ ,  $\|\cdot\|^2$ ,  $\mathbb{E}[\cdot]$  and  $\otimes$  denote the floor, Euclidean norm-square, expectation and Kronecker product operators, respectively.  $\text{sign}(\cdot)$  denotes the signum function and  $\text{sign}(0) = 1$  is assumed.  $[\mathbf{x}]_m$  denotes the  $m$ -th entry of a column vector  $\mathbf{x}$ .  $\mathbf{x}^{(m)}$  denotes the  $m$ -cyclic-down-shift version of  $\mathbf{x}$  for  $m > 0$  and  $|m|$ -cyclic-up-shift version of  $\mathbf{x}$  for  $m < 0$ .  $\text{diag}\{\mathbf{x}\}$  denotes a diagonal matrix with the elements of  $\mathbf{x}$  on its diagonal.  $[\mathbf{X}]_{m,n}$  denotes the  $(m,n)$ -th entry of a matrix  $\mathbf{X}$ .  $\mathbf{F}_N$  and  $\mathbf{I}_N$  denote the  $N \times N$  unitary DFT matrix and the  $N \times N$  identity matrix, respectively.  $\mathbf{e}_N^k$  denotes the  $k$ -th column vector of  $\mathbf{I}_N$ .  $\mathbf{1}_Q$  ( $\mathbf{0}_Q$ ) and  $\mathbf{0}_{P \times Q}$  denote the  $Q \times 1$  all-one (all-zero) vector and  $P \times Q$  all-zero matrix, respectively.  $\mathbf{J}_Q$  denotes the  $Q \times Q$  exchange matrix with ones on its anti-diagonal and zeros elsewhere. Unless otherwise stated,  $0 \leq \mu \leq N_t - 1$  and  $0 \leq \nu \leq N_r - 1$  are assumed.

## II. SYSTEM MODEL

Let us consider a MIMO-OFDM system with  $N_t$  transmit antennas,  $N_r$  receive antennas and  $N$  subcarriers. Suppose the training sequence transmitted from the  $\mu$ -th antenna is denoted by the  $N \times 1$  vector  $\tilde{\mathbf{t}}_\mu$ . Before transmission, this vector is processed by an inverse discrete Fourier transform (IDFT), and a cyclic prefix (CP) of length  $N_g$  is inserted. We assume that  $N_g \geq L - 1$ , where  $L$  is the maximum length of all the frequency selective fading channels. We further assume that all the transmit-receive antenna pairs are affected by the same

CFO. Define

$$\mathbf{D}_{\bar{N}}(\varepsilon) = \text{diag}\{[1, e^{j2\pi\varepsilon/N}, \dots, e^{j2\pi\varepsilon(\bar{N}-1)/N}]^T\},$$

where  $\varepsilon$  is the frequency offset normalized by the subcarrier spacing. Suppose the length- $L$  channel impulse response from the  $\mu$ -th transmit antenna to the  $\nu$ -th receive antenna is denoted by the  $L \times 1$  vector  $\mathbf{h}^{(\nu,\mu)}$ . Then, after removing the CP at the  $\nu$ -th receive antenna, the  $N \times 1$  received vector  $\mathbf{y}_\nu$  can be written as [12]

$$\mathbf{y}_\nu = \sqrt{N} e^{j2\pi\varepsilon N_g/N} \mathbf{D}_N(\varepsilon) \sum_{\mu=0}^{N_t-1} \left\{ \mathbf{F}_N^H \text{diag}\{\tilde{\mathbf{h}}^{(\nu,\mu)}\} \tilde{\mathbf{t}}_\mu \right\} + \mathbf{w}_\nu, \quad (1)$$

where

$$\tilde{\mathbf{h}}^{(\nu,\mu)} = \mathbf{F}_N [e_N^0, e_N^1, \dots, e_N^{L-1}] \mathbf{h}^{(\nu,\mu)},$$

and  $\mathbf{w}_\nu$  is an  $N \times 1$  vector of additive white complex Gaussian noise (AWGN) samples with zero-mean and equal variance of  $\sigma_w^2$ .

The goal of this paper is to design the training sequences  $\{\tilde{\mathbf{t}}_\mu\}_{\mu=0}^{N_t-1}$  and estimate  $\varepsilon$  from the observation of  $\{\mathbf{y}_\nu\}_{\nu=0}^{N_r-1}$ .

## III. FREQUENCY OFFSET ESTIMATION FOR MIMO OFDM

Let  $Q = N/P$  with  $(N)_P = 0$ . Design

$$(C0) \quad 0 \leq i_0 < i_1 < \dots < i_\mu < \dots < i_{N_t-1} < Q.$$

Define

$$\Theta_q = [e_N^q, e_N^{q+Q}, \dots, e_N^{(P-1)Q}], \quad 0 \leq q < Q.$$

Let  $\tilde{\mathbf{s}}_\mu$  denote a length- $P$  sequence whose elements are all non-zero. Then, we propose to construct the training sequence transmitted from the  $\mu$ -th antenna as  $\tilde{\mathbf{t}}_\mu = \Theta_{i_\mu} \tilde{\mathbf{s}}_\mu$  (C1). Without loss of generality, the total energy allocated to training is supposed to be split equally between the transmit antennas, i.e.,  $\|\tilde{\mathbf{s}}_\mu\|^2 = N/N_t$  (C2). Note that the entire training symbol is utilized by our proposed CFO estimator.

### A. ICFO Estimation

Let  $\mathbf{y} = [\mathbf{y}_0^T, \mathbf{y}_1^T, \dots, \mathbf{y}_\nu^T, \dots, \mathbf{y}_{N_r-1}^T]^T$  denote the  $N_r N \times 1$  cascaded vector from the  $N_r$  receive antennas. Then,  $\mathbf{y}$  can be written as

$$\mathbf{y} = \sqrt{N} e^{j2\pi\varepsilon N_g/N} \{\mathbf{I}_{N_r} \otimes [\mathbf{D}_N(\varepsilon) \mathbf{S}]\} \mathbf{h} + \mathbf{w}, \quad (2)$$

where

$$\begin{aligned} \mathbf{h} &= [\mathbf{h}_0^T, \mathbf{h}_1^T, \dots, \mathbf{h}_\nu^T, \dots, \mathbf{h}_{N_r-1}^T]^T, \\ \mathbf{h}_\nu &= [(\mathbf{h}^{(\nu,0)})^T, (\mathbf{h}^{(\nu,1)})^T, \dots, (\mathbf{h}^{(\nu,\mu)})^T, \dots, (\mathbf{h}^{(\nu,N_t-1)})^T]^T, \\ \mathbf{S} &= \bar{\mathbf{F}}^H \text{diag}\{[\tilde{\mathbf{s}}_0^T, \tilde{\mathbf{s}}_1^T, \dots, \tilde{\mathbf{s}}_\mu^T, \dots, \tilde{\mathbf{s}}_{N_t-1}^T]^T\} \bar{\mathbf{F}}, \\ \bar{\mathbf{F}} &= [\Theta_{i_0}, \Theta_{i_1}, \dots, \Theta_{i_\mu}, \dots, \Theta_{i_{N_t-1}}]^T \mathbf{F}_N, \\ \bar{\mathbf{F}} &= [e_{N_t}^0 \otimes \Theta_{i_0}^T, e_{N_t}^1 \otimes \Theta_{i_1}^T, \dots, e_{N_t}^{\mu} \otimes \Theta_{i_\mu}^T, \\ &\quad \dots, e_{N_t}^{N_t-1} \otimes \Theta_{i_{N_t-1}}^T] \{\mathbf{I}_{N_t} \otimes [\mathbf{F}_N [\mathbf{I}_L, \mathbf{0}_{L \times (N-L)}]^T]\}, \\ \mathbf{w} &= [\mathbf{w}_0^T, \mathbf{w}_1^T, \dots, \mathbf{w}_\nu^T, \dots, \mathbf{w}_{N_r-1}^T]^T. \end{aligned}$$

From (2), the ML estimation of  $\varepsilon$  can be readily obtained [20] as follows

$$\hat{\varepsilon} = \arg \max_{\tilde{\varepsilon}} \{ \mathbf{y}^H \{ \mathbf{I}_{N_r} \otimes [\mathbf{D}_N(\tilde{\varepsilon}) \tilde{\mathbf{S}} (\tilde{\mathbf{S}}^H \tilde{\mathbf{S}})^{-1} \tilde{\mathbf{S}}^H \mathbf{D}_N(-\tilde{\varepsilon})] \} \mathbf{y} \}, \quad (3)$$

where

$$\tilde{\mathbf{S}} = \bar{\mathbf{F}}^H \text{diag}\{[\tilde{\mathbf{s}}_0^T, \tilde{\mathbf{s}}_1^T, \dots, \tilde{\mathbf{s}}_\mu^T, \dots, \tilde{\mathbf{s}}_{N_t-1}^T]^T\}.$$

With condition (C0), which implies the orthogonality of the training sequences in the frequency domain, matrix inversion involved in (3) can be avoided, and (3) can thus be simplified as follows

$$\hat{\varepsilon} = \arg \max_{\tilde{\varepsilon}} \{ \| \{ \mathbf{I}_{N_r} \otimes [\bar{\mathbf{F}} \mathbf{D}_N(-\tilde{\varepsilon})] \} \mathbf{y} \|^2 \}. \quad (4)$$

Actually, with condition (C0), our proposed training sequences can be treated as frequency-division multiplexing (FDM) pilot allocation type sequences [18].

Let  $\mathbf{l}$  denote the pilot location vector which is given by  $\mathbf{l} = \sum_{\mu=0}^{N_t-1} e^{j\mu} \mathbf{e}^{\mu}$ . Then, we have as follows.

*Theorem 1:* With  $\tilde{\varepsilon}, \varepsilon \in (-[Q/2], Q - [Q/2])$ ,  $\tilde{\varepsilon} = \varepsilon$  uniquely maximizes the cost function in (4) for any  $\mathbf{h}^{(\nu, \mu)} (\neq \mathbf{0}_L)$  with the following condition

$$(C3) \quad (N - N_t P) \geq N_t P \ \& \ P \geq L \ \& \quad (\mathbf{1}_Q - \mathbf{l})^T \mathbf{l}^{(q)} > 0, \forall q \in \{1, 2, \dots, Q-1\}.$$

*Proof:* See Appendix I. ■

It follows immediately that the corresponding estimation is identifiable for  $\varepsilon \in (-[Q/2], Q - [Q/2])$  if condition (C3) is satisfied.

From (4), the CFO can be estimated by exploiting fast Fourier transform (FFT) interpolation and a time consuming line search over a large set of frequency grids. But this approach is computationally complicated and the estimate precision also depends on the FFT size used. To reduce complexity,  $\varepsilon$  is divided into the ICFO  $\varepsilon_i$  and the FCFO  $\varepsilon_f$ . The ICFO  $\varepsilon_i$  can be estimated by invoking only the  $N$  point FFT as follows

$$\hat{\varepsilon}_i = \arg \max_{-\lfloor Q/2 \rfloor < \tilde{\varepsilon}_i \leq Q - \lfloor Q/2 \rfloor} \{ \| \{ \mathbf{I}_{N_r} \otimes [\bar{\mathbf{F}} \mathbf{D}_N(-\tilde{\varepsilon}_i)] \} \mathbf{y} \|^2 \}. \quad (5)$$

The discussion on the uniqueness of  $\hat{\varepsilon}_i = \varepsilon_i$  in maximizing the estimation metric for any  $\mathbf{h}^{(\nu, \mu)} (\neq \mathbf{0}_L)$  is provided in Appendix II. Other ICFO estimators may also be used. Once the ICFO is estimated, ICFO correction can be readily carried out as follows

$$\bar{\mathbf{y}} = e^{-j2\pi\hat{\varepsilon}_i N_g/N} [\mathbf{I}_{N_r} \otimes \mathbf{D}_N(-\hat{\varepsilon}_i)] \mathbf{y}. \quad (6)$$

### B. FCFO Estimation

The FCFO  $\varepsilon_f$  is estimated based on  $\bar{\mathbf{y}}$ . Assume  $\hat{\varepsilon}_i = \varepsilon_i$  and substitute (2) into (6). Then, by exploiting condition (C0)

and condition (C1), which implies the periodic property of the training sequences,  $\bar{\mathbf{y}}$  can be expressed in an equivalent form as shown in (7) at the bottom of the page. In (7),

$$\beta_\mu = \varepsilon_f + i_\mu, \\ \mathbf{v} = e^{-j2\pi\varepsilon_i N_g/N} [\mathbf{I}_{N_r} \otimes \mathbf{D}_N(-\varepsilon_i)] \mathbf{w}.$$

It follows from (7) that the estimation of  $\varepsilon_f$  is equivalent to the estimation of the  $N_t$  different equivalent CFOs  $\{\beta_\mu\}_{\mu=0}^{N_t-1}$ .

Furthermore, exploiting the periodic property of the training sequences again, we can stack  $\bar{\mathbf{y}}$  into the  $Q \times N_r P$  matrix  $\mathbf{Y} = [\mathbf{Y}_0, \mathbf{Y}_1, \dots, \mathbf{Y}_\nu, \dots, \mathbf{Y}_{N_r-1}]$ , where

$$[\mathbf{Y}_\nu]_{q,p} = [((e^{j\nu})^T \otimes \mathbf{I}_N) \bar{\mathbf{y}}]_{qP+p}, \quad 0 \leq q < Q, 0 \leq p < P.$$

Then, (7) can be expressed in the following equivalent form

$$\mathbf{Y} = \mathbf{B} \mathbf{X} + \mathbf{V}, \quad (8)$$

where

$$\mathbf{B} = [\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_\mu, \dots, \mathbf{b}_{N_t-1}], \\ \mathbf{b}_\mu = [1, e^{j2\pi\beta_\mu/Q}, \dots, e^{j2\pi\beta_\mu q/Q}, \dots, e^{j2\pi\beta_\mu(Q-1)/Q}]^T, \\ \mathbf{X} = [\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_\nu, \dots, \mathbf{X}_{N_r-1}], \\ \mathbf{X}_\nu = [\mathbf{x}^{(\nu,0)}, \mathbf{x}^{(\nu,1)}, \dots, \mathbf{x}^{(\nu,\mu)}, \dots, \mathbf{x}^{(\nu,N_t-1)}]^T, \\ \mathbf{x}^{(\nu,\mu)} = \sqrt{P} e^{j2\pi\varepsilon_f N_g/N} \mathbf{D}_P(\beta_\mu) \\ \times \mathbf{F}_P^H \text{diag}\{\tilde{\mathbf{s}}_\mu\} \Theta_{i_\mu}^T \mathbf{F}_N [\mathbf{I}_L, \mathbf{0}_{L \times (N-L)}]^T \mathbf{h}^{(\nu,\mu)},$$

and  $\mathbf{V}$  is the  $Q \times N_r P$  matrix generated from  $\mathbf{v}$  in the same way as  $\mathbf{Y}$ . From (8), we can see that there exists an inherent relationship between FCFO estimation and DOA estimation in ULA [21] with  $P$ ,  $Q$  and  $N_r$  satisfying  $N_r P > Q$  (C4).

The covariance matrix of  $\mathbf{Y}$  can be estimated by  $\hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}} = \mathbf{Y}\mathbf{Y}^H / (N_r P)$ . Let  $\mathbf{L}$  denote the  $Q \times Q$  unitary column conjugate symmetric matrix which is given by

$$\mathbf{L} = \frac{[1 - (Q)2]}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_{Q/2} & j\mathbf{I}_{Q/2} \\ \mathbf{J}_{Q/2} & -j\mathbf{J}_{Q/2} \end{bmatrix} \\ + \frac{(Q)2}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_{(Q-1)/2} & \mathbf{0}_{(Q-1)/2} & j\mathbf{I}_{(Q-1)/2} \\ \mathbf{0}_{(Q-1)/2}^T & \sqrt{2} & \mathbf{0}_{(Q-1)/2}^T \\ \mathbf{J}_{(Q-1)/2} & \mathbf{0}_{(Q-1)/2} & -j\mathbf{J}_{(Q-1)/2} \end{bmatrix}.$$

Then, with the aid of  $\mathbf{L}$ , the complex matrix  $\hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}}$  can be transformed into a real matrix [22] as follows

$$\hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}}^r = 1/2 \cdot \mathbf{L}^H (\hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}} + \mathbf{J}_Q \hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}}^* \mathbf{J}_Q) \mathbf{L} \\ = \Re(\mathbf{L}^H \hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}} \mathbf{L}). \quad (9)$$

The eigen-decomposition of the real matrix  $\hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}}^r$  can be obtained as

$$\hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}}^r = \mathbf{E}_X \mathbf{\Lambda}_X \mathbf{E}_X^H + \mathbf{E}_V \mathbf{\Lambda}_V \mathbf{E}_V^H, \quad (10)$$

where

$$\mathbf{\Lambda}_X = \text{diag}\{[\lambda_0, \lambda_1, \dots, \lambda_{N_t-1}]^T\},$$

$$\bar{\mathbf{y}} = \sqrt{N} e^{j2\pi\varepsilon_f N_g/N} \{ \mathbf{I}_{N_r} \otimes \{ [\mathbf{D}_N(\beta_0), \mathbf{D}_N(\beta_1), \dots, \mathbf{D}_N(\beta_\mu), \dots, \mathbf{D}_N(\beta_{N_t-1})] \} \\ \times (\mathbf{I}_{N_t} \otimes \mathbf{1}_Q \otimes \mathbf{F}_P^H) \text{diag}\{[\tilde{\mathbf{s}}_0^T, \tilde{\mathbf{s}}_1^T, \dots, \tilde{\mathbf{s}}_\mu^T, \dots, \tilde{\mathbf{s}}_{N_t-1}^T]^T\} \check{\mathbf{F}} \} \mathbf{h} + \mathbf{v}, \quad (7)$$

$$\Lambda_{\mathbf{V}} = \sigma_w^2 \mathbf{I}_{Q-N_t},$$

$$\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{N_t-1} > \sigma_w^2,$$

and  $\mathbf{E}_{\mathbf{X}}$  and  $\mathbf{E}_{\mathbf{V}}$  contain the unitary eigen-vectors that span the signal space and noise space, respectively. Let

$$z = e^{j2\pi\beta/Q},$$

$$\mathbf{a}(z) = [1, z, \dots, z^{Q-1}]^T.$$

Then, by exploiting  $\mathbf{a}(z)$ , a polynomial of degree  $2(Q-1)$  with complex coefficients can be obtained [23] as follows

$$f(z) = \mathbf{a}^T(z) \mathbf{J}_Q \mathbf{A} \mathbf{a}(z), \quad (11)$$

where

$$\mathbf{A} = \mathbf{L}[\mathbf{I}_Q - \mathbf{E}_{\mathbf{X}} \mathbf{E}_{\mathbf{X}}^H] \mathbf{L}^H.$$

Correspondingly,  $\{\beta_\mu\}_{\mu=0}^{N_t-1}$  can be indirectly estimated by calculating the pairwise roots of  $f(z) = 0$  which are closest to the unit circle.

Note that by its definition,  $z$  is always located on the unit circle. By analysis, we find that the following geometric mapping exists,

$$g(z) = \cot(\pi\beta/Q) = j(z+1)/(z-1). \quad (12)$$

It follows immediately that  $g(z)$  is a monotonic reversible mapping with real values for  $\beta \in [-0.5, Q-0.5)$ . Hence, we have  $z(g) = (g+j)/(g-j)$ . Then,  $\mathbf{a}(z)$  can be expressed as a function of  $g$  as

$$\mathbf{a}(z) = (g-j)^{1-Q} \mathbf{d}(g), \quad (13)$$

where

$$\mathbf{d}(g) = [(g-j)^{Q-1}, (g+j)(g-j)^{Q-2}, \dots, (g+j)^{Q-1}]^T.$$

Since the elements of  $\mathbf{d}(g)$  are polynomials of degree  $(Q-1)$  with respect to  $g$ ,  $\mathbf{d}(g)$  in (13) can be further expressed as

$$\mathbf{d}(g) = \Phi \mathbf{a}(g), \quad (14)$$

where  $\Phi$  is the  $Q \times Q$  coefficient matrix. By straight-forward calculation, the elements of  $\Phi$  are obtained as shown in (15). In (15),  $C_q^{q''} = q! / [(q-q'')! q''!]$ . Then, from (15), we immediately obtain

$$[\Phi]_{q,q'} = [\Phi]_{Q-1-q,q'}^* \ \& \ (Q)_2 \Im\{[\Phi]_{(Q-1)/2,q}\} = 0,$$

$$\forall q, q' \in \{0, 1, \dots, Q-1\}, \quad (16)$$

which shows that  $\Phi$  is a column conjugate symmetric matrix *no matter*  $Q$  is even or odd.

By exploiting (13) and (14) in (11), the following equivalent polynomial can be obtained (ignoring the constant items)

$$f^r(g) = \mathbf{a}^T(g) \mathbf{J}_Q \mathbf{A}^r \mathbf{a}(g), \quad (17)$$

where

$$\mathbf{A}^r = \mathbf{J}_Q \Phi^H \mathbf{L} [\mathbf{I}_Q - \mathbf{E}_{\mathbf{X}} \mathbf{E}_{\mathbf{X}}^H] \mathbf{L}^H \Phi.$$

Due to the column conjugate symmetric property of  $\mathbf{L}$  and  $\Phi$ , we have (18) as shown at the bottom of the page. Hence,  $\Phi^H \mathbf{L}$  becomes a real matrix *no matter*  $Q$  is even or odd. Moreover,  $\mathbf{E}_{\mathbf{X}}$  is also a real matrix. Accordingly,  $f^r(g)$  is transformed into a polynomial with real coefficients. The roots of  $f^r(g) = 0$  can be obtained by the fast root-calculating algorithms for real polynomials in [24], whose computational complexities are much less than those for complex ones.

After the roots of  $f^r(g) = 0$  are obtained, the FCFO can be readily estimated according to the following steps.

- 1) Find the  $N_t$  pairwise roots of  $f^r(g) = 0$  whose imaginary parts have the smallest absolute values,  $\{\Re(g_\mu) \pm j\Im(g_\mu)\}_{\mu=0}^{N_t-1}$ .
- 2) Calculate the equivalent CFOs corresponding to the  $N_t$  roots,

$$\hat{\beta}_\mu = (Q/\pi \times \text{acot}[\Re(g_\mu)])_Q. \quad (19)$$

- 3) Calculate the FCFO  $\hat{\epsilon}_\mu^f$  corresponding to each  $\hat{\beta}_\mu$ ,

$$\hat{\epsilon}_\mu^f = \begin{cases} \hat{\beta}_\mu - i_0, & \hat{\beta}_\mu \in ((i_0 - \epsilon_{\text{th}})_Q, i_0 + \epsilon_{\text{th}}) \\ \hat{\beta}_\mu - i_1, & \hat{\beta}_\mu \in (i_1 - \epsilon_{\text{th}}, i_1 + \epsilon_{\text{th}}) \\ \dots & \dots \\ \hat{\beta}_\mu - i_{N_t-1}, & \hat{\beta}_\mu \in (i_{N_t-1} - \epsilon_{\text{th}}, (i_{N_t-1} + \epsilon_{\text{th}})_Q) \end{cases}, \quad (20)$$

where  $\epsilon_{\text{th}}$  denotes a threshold which is predefined to avoid the ambiguous estimation and  $0.5 < \epsilon_{\text{th}} < 1$ ,  $[a, b)$  has the usual meaning for  $a < b$ , and  $[a, b) = [0, b) \cup [a, Q)$  for  $a > b$ .

- 4) Obtain the final FCFO by averaging  $\{\hat{\epsilon}_\mu^f\}_{\mu=0}^{N_t-1}$ ,

$$\hat{\epsilon}_f = \frac{1}{N_t} \sum_{\mu=0}^{N_t-1} \hat{\epsilon}_\mu^f. \quad (21)$$

$$[\Phi]_{q,q'} = j^{Q-1-q'} \sum_{q''=\max\{0, q+q'-Q+1\}}^{\min\{q,q'\}} \left\{ C_q^{q''} C_{Q-1-q}^{q'-q''} (-1)^{Q-1-q-q'+q''} \right\}, \quad 0 \leq q, q' \leq Q-1, \quad (15)$$

$$[\Phi^H \mathbf{L}]_{q,q'} = \sum_{q''=0}^{[1-(Q)_2]Q/2+(Q)_2(Q-1)/2-1} \left\{ [\Phi]_{q'',q}^* [\mathbf{L}]_{q'',q'} + [\Phi]_{q'',q} [\mathbf{L}]_{q'',q'}^* \right\} + (Q)_2 [\Phi]_{(Q-1)/2,q} [\mathbf{L}]_{(Q-1)/2,q'}$$

$$= 2\Re \left\{ \sum_{q''=0}^{[1-(Q)_2]Q/2+(Q)_2(Q-1)/2-1} \left\{ [\Phi]_{q'',q}^* [\mathbf{L}]_{q'',q'} \right\} \right\} + (Q)_2 [\Phi]_{(Q-1)/2,q} [\mathbf{L}]_{(Q-1)/2,q'},$$

$$\forall q, q' \in \{0, 1, \dots, Q-1\}. \quad (18)$$

### C. Computational Complexity

The computational load of our proposed CFO estimator mainly involves the  $N$  point FFT, the eigen-decomposition of  $\hat{\mathbf{R}}_{\mathbf{Y}\mathbf{Y}}^r$  and the root-calculation for  $f^r(g) = 0$ , which require  $4N \log_2 N$ ,  $9Q^3$  and  $64/3 \cdot (Q-1)^3$  real additions or multiplications [24] [25], respectively. Compared with the direct CFO estimate from (4), which requires a large point DFT operation and an exhaustive line search, our proposed estimator has significantly lower complexity especially with a relatively small  $Q$ . Furthermore, by applying the approach in [26], the complexity of our proposed estimator can be further decreased by calculating the roots from the first-order derivative of  $f^r(g) = 0$ , but at the cost of a slight performance degradation at low signal-to-noise ratio (SNR).

### IV. TRAINING SEQUENCE DESIGN

With our design conditions in the previous section, the training sequences are determined completely by the base sequences  $\{\tilde{\mathbf{s}}_\mu\}_{\mu=0}^{N_t-1}$  with fixed  $P$ ,  $Q$  and  $\{i_\mu\}_{\mu=0}^{N_t-1}$ . Let  $\mathbf{R}_{\mathbf{X}\mathbf{X}}$  denote the covariance matrix of  $\mathbf{X}$ . It is pointed out in [21] [22] that the optimal performance can be achieved with uncorrelated signals, i.e., diagonal matrix  $\mathbf{R}_{\mathbf{X}\mathbf{X}}$ . Since  $\mathbf{R}_{\mathbf{X}\mathbf{X}}$  can be estimated by  $\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}} = \mathbf{X}\mathbf{X}^H/(N_r P)$ , it is expected that good performance can be achieved with diagonal matrix  $\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}$ . Making  $\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}$  to be a diagonal matrix is equivalent to making  $[\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}]_{\mu,\mu'}|_{\mu \neq \mu'} = 0$ . Define

$$\begin{aligned} \varpi_{\mu,\mu'} &= (i_\mu - i_{\mu'})/Q, \\ \mathbf{s}_\mu &= \frac{1}{\sqrt{Q}} \mathbf{F}_P^H \tilde{\mathbf{s}}_\mu. \end{aligned}$$

Assume that the channel taps remain constant during the training period. Then,  $[\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}]_{\mu,\mu'}$  can be expressed as

$$[\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}]_{\mu,\mu'} = \frac{Q}{N_r} \sum_{\nu=0}^{N_r-1} \sum_{l=0}^{L-1} \sum_{l'=0}^{L-1} \left\{ [\mathbf{h}^{(\nu,\mu)}]_l [\mathbf{T}^{(\mu,\mu')}]_{l,l'} [\mathbf{h}^{(\nu,\mu')}]_{l'}^* \right\}, \quad (22)$$

where

$$[\mathbf{T}^{(\mu,\mu')}]_{l,l'} = (\mathbf{s}_\mu^{(l)})^T \mathbf{D}_P(\varpi_{\mu,\mu'} Q) (\mathbf{s}_{\mu'}^{(l')})^*.$$

Since the elements of  $\mathbf{h}^{(\nu,\mu)}$  are random variables,  $[\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}]_{\mu,\mu'}|_{\mu \neq \mu'} = 0$  can be achieved by making the training sequences satisfy

$$[\mathbf{T}^{(\mu,\mu')}]_{l,l'} = 0, \text{ if } \mu \neq \mu' \text{ \& } 0 \leq l, l' \leq L-1. \quad (23)$$

It follows immediately that the optimal training sequences should satisfy the condition in (23). However, with condition (C0),  $\varpi_{\mu,\mu'}$  is definitely a decimal fraction, which greatly complicates the satisfaction of the condition in (23) with proper training sequences. To ease the above problem, we construct two types of sub-optimal training sequences which can make the off-diagonal elements of  $\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}$  as small as possible. It has been proven in [15] that training sequence with the zero auto-correlation (ZAC) property is optimal for CFO estimation in SISO frequency selective fading channels. Besides, constant amplitude ZAC (CAZAC) sequence (e.g., [19], [27] and references therein) is often a preferred choice for training. Therefore, we propose to construct the training

sequences from a length- $P$  Chu sequence  $\mathbf{s}$  with its element given by

$$[\mathbf{s}]_p = e^{j\pi v p^2/P}, \quad 0 \leq p \leq P-1,$$

where  $v$  is coprime to  $P$ , and  $P$  is supposed to be even.

Let  $\tilde{\mathbf{s}}_\mu = \sqrt{Q/N_t} \mathbf{F}_P \mathbf{s}^{(\mu M)}$  with  $M = \lfloor P/N_t \rfloor$  and  $N_t \geq N_t$ . Then, we refer to the so-constructed training sequences as TS 0. Let  $\tilde{\mathbf{s}}_\mu = \sqrt{Q/N_t} \mathbf{F}_P \mathbf{s}$ . Then, we refer to the so-constructed training sequences as TS 1. Note that the TS 1 training sequences are equivalent to the so-called repeated phase-rotated Chu (RPC) sequences [28]. Define

$$p_{\mu,l} = (1-m)\mu M + l,$$

where  $m = 0$  for TS 0, and  $m = 1$  for TS 1. Then, from the above constructions we have

$$\begin{aligned} [\mathbf{T}^{(\mu,\mu')}]_{l,l'} &= \frac{1}{N_t} (-1)^{v(p_{\mu,l} - p_{\mu',l'}) + 1} e^{j\pi v(p_{\mu,l}^2 - p_{\mu',l'}^2)/P} \\ &\quad \times e^{-j\pi(P-1)[v(p_{\mu,l} - p_{\mu',l'}) - \varpi_{\mu,\mu'}]/P} \sin(\pi \varpi_{\mu,\mu'}) \\ &\quad / \sin\{\pi[v(p_{\mu,l} - p_{\mu',l'}) - \varpi_{\mu,\mu'}]/P\}. \end{aligned} \quad (24)$$

It follows immediately from (24) that

$$\begin{aligned} |[\mathbf{T}^{(\mu,\mu')}]_{l,l'}|_{p_{\mu,l} - p_{\mu',l'} \neq 0} \\ \ll |[\mathbf{T}^{(\mu,\mu')}]_{l,l'}|_{p_{\mu,l} - p_{\mu',l'} = 0} < P/N_t. \end{aligned} \quad (25)$$

We see that  $|[\mathbf{T}^{(\mu,\mu')}]_{l,l'}|$  achieves its maximum for TS 0 when  $(\mu - \mu')M = l' - l$  and for TS 1 when  $l = l'$ . Assume that the channel energy is mainly concentrated in the preceding  $M$  channel taps and the first channel tap is the dominant one. Then, it can be inferred from (22) and (25) that the value of  $|[\hat{\mathbf{R}}_{\mathbf{X}\mathbf{X}}]_{\mu,\mu'}|_{\mu \neq \mu'}$  for TS 0 is very small, and it is much smaller than that for TS 1 in the same channel environment. In this sense, we can say that TS 0 is superior to TS 1, which will be verified through simulation results in the following section.

Note that our proposed training sequence structure is similar to the one introduced recently in [16], and the identifiability of our CFO estimator, however, cannot be guaranteed with the training sequences in [16].

### V. SIMULATION RESULTS

To evaluate our CFO estimator's performance with the proposed training sequences for MIMO-OFDM, a number of simulations are carried out. Throughout the simulations, a MIMO OFDM system of bandwidth 20MHz operating at 5GHz with  $N = 1024$  and  $N_g = 64$  is used. Each channel has 4 independent Rayleigh fading taps, whose relative average-powers and propagation delays are  $\{0, -9.7, -19.2, -22.8\}$ dB and  $\{0, 0.1, 0.2, 0.4\}$ μs, respectively. For the training sequences, we set  $P$  and  $Q$  to 64 and 16, respectively. The normalized CFO  $\varepsilon$  is generated within the range  $(-Q/2, Q - Q/2]$ . For description convenience, we henceforth refer to the proposed real polynomial based CFO estimator as RPBE.

In the following, we use the average CRB (avCRB), which corresponds to the extended Miller and Chang bound (EMCB) [15] [29], to benchmark the performance of RPBE. The average CRB or EMCB is obtained by simply averaging the snapshot CRB over independent channel realizations, which

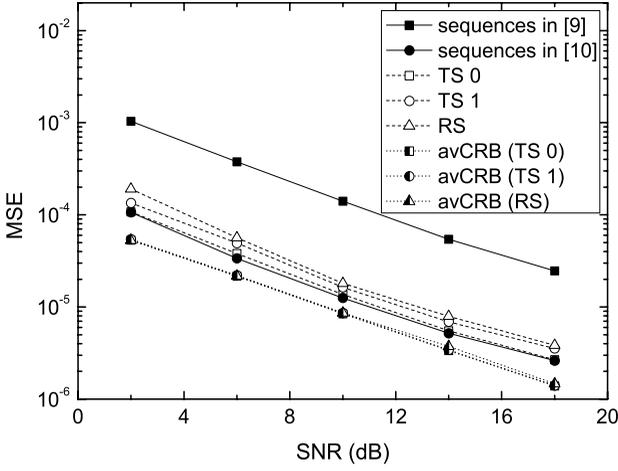


Fig. 1. CFO estimation performance for different training sequences with  $N_t = 3$  and  $N_r = 2$ .

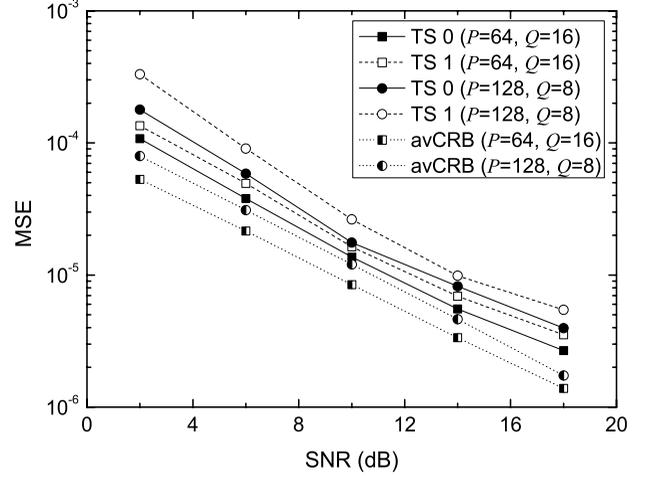


Fig. 2. CFO estimation performance for different  $Q$  with  $N_t = 3$  and  $N_r = 2$ .

can be calculated straight-forwardly [20] as follows

$$\text{CRB}_\varepsilon = \frac{N\sigma_w^2}{8\pi^2 \mathbf{h}^H \boldsymbol{\mathcal{X}}^H \mathbf{B} [\mathbf{I}_{N_r N} - \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}^H \boldsymbol{\mathcal{X}})^{-1} \boldsymbol{\mathcal{X}}^H] \mathbf{B} \boldsymbol{\mathcal{X}} \mathbf{h}}, \quad (26)$$

where

$$\begin{aligned} \boldsymbol{\mathcal{X}} &= \mathbf{I}_{N_r} \otimes \mathbf{S}, \\ \mathbf{B} &= \mathbf{I}_{N_r} \otimes \text{diag}\{[N_g, N_g + 1, \dots, N_g + N - 1]^T\}. \end{aligned}$$

In order to verify our analysis concerning training sequence design with the Chu sequence, we construct the random sequences (RS) by generating the  $N_t P$  pilots randomly, and compare the MSE performances of RPBE with TS 0, TS 1 and RS. Also included for comparison are the performances of the CFO estimators with their training sequences in [9] and [10]. In Fig. 1, we present the corresponding simulation results with  $N_t = 3$  and  $N_r = 2$ . It can be observed that the performance of RPBE with TS 0 is slightly better than that with TS 1, which coincides with the analytical results concerning training sequence design in the previous section. It can also be observed that the performances of RPBE with TS 0 and TS 1 are better than that with RS, which should be attributed to the good correlation property of TS 0 and TS 1. Actually, the performance improvements are more evident when the frequency selective fading channels with large delay spreads are considered. Furthermore, we observe that the performances of RPBE with TS 0 and TS 1 are far better than that of the CFO estimator in [9], and almost the same as that of the CFO estimator in [10]. Noting that the overhead of the training sequences presented in [10] grows linearly with the number of transmit antennas, we can find certain advantages of the proposed training sequences. We also observe that the average CRB for TS 0 is slightly smaller than that for TS 1. Since the average CRBs for TS 0 and TS 1 are very close, only one curve is plotted subsequently.

Depicted in Fig. 2 is the MSE performance of RPBE as a function of SNR for different  $Q$  with  $N_t = 3$  and  $N_r = 2$ .

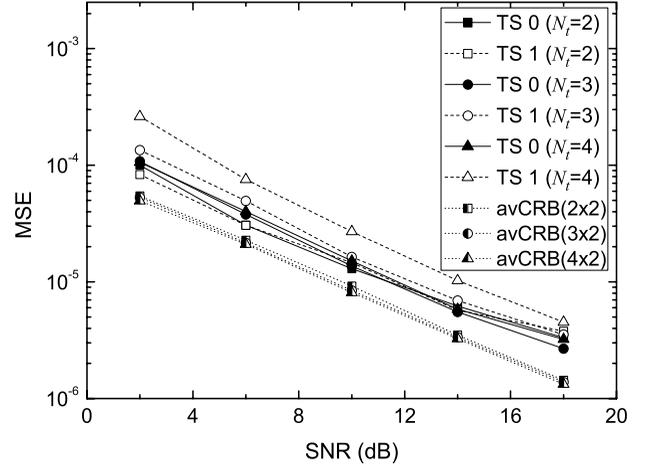


Fig. 3. CFO estimation performance for different numbers of transmit antennas with  $N_r = 2$ .

Studying the curves in Fig. 2, we can see that the performance of RPBE with the same training sequences degrades with smaller  $Q$  provided that  $Q > N_t$  and  $P \geq L$ , which agrees well with the analysis in [21]. We can also see that a smaller  $Q$  yields a larger average CRB, and the MSE performance follows the same trend as the average CRB.

In Fig. 3, we illustrate how the number of transmit antennas  $N_t$  affects the performance of RPBE. It can be observed that the MSE performance for TS 1 deteriorates with increased  $N_t$  due to the influence of MAI, whereas the MSE performance for TS 0 degrades slightly, which should be ascribed to the better correlation property of TS 0. Another observation is that the number of transmit antennas has little impact on the average CRB.

Fig. 4 shows how the number of receive antennas  $N_r$  affects the performance of RPBE. We can observe that the MSE performance of RPBE is substantially improved for both TS 0 and TS 1 by increasing  $N_r$ . We have the similar observation for the average CRB. These observations imply that it is more

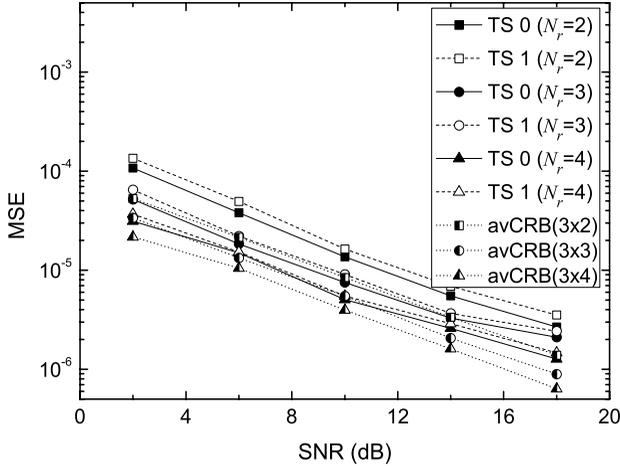


Fig. 4. CFO estimation performance for different numbers of receive antennas with  $N_t = 3$ .

efficient for performance improvement of RPBE to increase the number of receive antennas than the number of transmit antennas.

## VI. CONCLUSIONS

In this paper, we have presented a training sequence assisted CFO estimator for MIMO OFDM systems by exploiting the properties of the training sequences and an efficient geometric mapping. We have developed the required conditions for the training sequences to yield estimation identifiability, and proposed sub-optimal training sequences constructed from the Chu sequence. Our proposed estimator and training sequences yield better or similar estimation performance with much smaller training overhead than existing methods for MIMO-OFDM systems.

## APPENDIX I

This appendix presents the proof of Theorem 1. Substitute (2) into the cost function in (4) and ignore the noise item. Define  $\Delta = \varepsilon - \tilde{\varepsilon}$ . Then, the cost function in (4) can be expressed as

$$\mathfrak{P}(\Delta) = N\mathbf{h}^H \{ \mathbf{I}_{N_r} \otimes [\mathbf{S}^H \mathbf{D}_N(-\Delta) \check{\mathbf{F}}^H \check{\mathbf{F}} \mathbf{D}_N(\Delta) \mathbf{S}] \} \mathbf{h}. \quad (27)$$

Let

$$\check{\mathbf{S}} = \text{diag}\{[\check{\mathbf{s}}_0^T, \check{\mathbf{s}}_1^T, \dots, \check{\mathbf{s}}_\mu^T, \dots, \check{\mathbf{s}}_{N_t-1}^T]^T\}.$$

Define

$$\mathfrak{D}(\Delta) = N\mathbf{h}^H [\mathbf{I}_{N_r} \otimes (\check{\mathbf{F}}^H \check{\mathbf{S}}^H \check{\mathbf{S}} \check{\mathbf{F}})] \mathbf{h} - \mathfrak{P}(\Delta).$$

Then, it follows immediately that the maximum of  $\mathfrak{P}(\Delta)$  corresponds to the minimum of  $\mathfrak{D}(\Delta)$  with respect to  $\Delta$ . Suppose

$$0 \leq \omega < Q - N_t \ \& \ 0 \leq z_\omega < Q \ \& \ z_\omega \neq i_\mu \ \& \\ z_\omega = z_{\omega'}, \ \text{iff.} \ \omega = \omega'.$$

Then,  $\mathfrak{D}(\Delta)$  can be expressed as

$$\mathfrak{D}(\Delta) = \mathbf{d}^H(\Delta) \mathbf{d}(\Delta), \quad (28)$$

where

$$\mathbf{d}(\Delta) = [\mathbf{d}_0^T(\Delta), \mathbf{d}_1^T(\Delta), \dots, \mathbf{d}_\nu^T(\Delta), \dots, \mathbf{d}_{N_r-1}^T(\Delta)]^T,$$

$$\mathbf{d}_\nu(\Delta) = \mathbf{C} \check{\mathbf{S}} \check{\mathbf{F}} \mathbf{h}_\nu,$$

$$\mathbf{C} = \begin{bmatrix} \mathbf{c}^{(0,0)} & \dots & \mathbf{c}^{(0,N_t-1)} \\ \vdots & \mathbf{c}^{(\omega,\mu)} & \vdots \\ \mathbf{c}^{(Q-N_t-1,0)} & \dots & \mathbf{c}^{(Q-N_t-1,N_t-1)} \end{bmatrix},$$

$$\mathbf{c}^{(\omega,\mu)} = \sqrt{N} \Theta_{z_\omega}^T \mathbf{F}_N \mathbf{D}_N(\Delta) \mathbf{F}_N^H \Theta_{i_\mu}.$$

Correspondingly, we have

$$\mathfrak{D}(\Delta) \geq 0 \ \& \ \mathfrak{D}(0) = 0 \ \&$$

$$\mathfrak{D}(\Delta) = 0, \ \text{iff.} \ \mathbf{d}(\Delta) = \mathbf{0}_{N_r(N-N_tP)}. \quad (29)$$

Therefore, for any  $\mathbf{h}^{(\nu,\mu)} (\neq \mathbf{0}_L)$ , proving that  $\tilde{\varepsilon} = \varepsilon$  is the unique value to maximize  $\mathfrak{P}(\Delta)$  with  $\tilde{\varepsilon}, \varepsilon \in (-\lfloor Q/2 \rfloor, Q - \lfloor Q/2 \rfloor]$  is equivalent to proving that  $\mathbf{d}(\Delta) \neq \mathbf{0}_{N_r(N-N_tP)}$  for any  $\Delta \in (-Q, 0) \cup (0, Q)$ .

Consider the following two cases:

1) When  $\Delta$  is not an integer: We establish immediately from the definition of  $\mathbf{C}^{(\omega,\mu)}$  that it is a column-wise circulant matrix. Denote the square matrix formed by the first  $N_t P$  rows of  $\mathbf{C}$  by  $\mathbf{C}_P$  (recall that  $(N - N_t P) \geq N_t P$  in our design condition). Then, by exploiting the relation between circulant matrix and DFT matrix [25],  $\mathbf{C}_P$  can be decomposed as

$$\mathbf{C}_P = \sqrt{P} (\mathbf{I}_{N_t} \otimes \mathbf{F}_P^H) \mathbf{\Lambda} (\mathbf{I}_{N_t} \otimes \mathbf{F}_P), \quad (30)$$

where

$$\mathbf{\Lambda} = \begin{bmatrix} \mathbf{\Lambda}^{(0,0)} & \dots & \mathbf{\Lambda}^{(0,N_t-1)} \\ \vdots & \mathbf{\Lambda}^{(\mu,\mu')} & \vdots \\ \mathbf{\Lambda}^{(N_t-1,0)} & \dots & \mathbf{\Lambda}^{(N_t-1,N_t-1)} \end{bmatrix},$$

$$\mathbf{\Lambda}^{(\mu,\mu')} = \text{diag}\{\tilde{\mathbf{c}}^{(\mu,\mu')}\},$$

$$\tilde{\mathbf{c}}^{(\mu,\mu')} = \mathbf{F}_P \mathbf{C}^{(\mu,\mu')} \mathbf{e}_P^0.$$

The element of  $\tilde{\mathbf{c}}^{(\mu,\mu')}$  can be obtained from its definition as follows

$$[\tilde{\mathbf{c}}^{(\mu,\mu')}]_p = \frac{1}{\sqrt{Q}} c^{(\mu,\mu')} (1 - e^{j2\pi\Delta}) e^{j2\pi(i_{\mu'} - z_\mu + \Delta)(-p)P/N} \neq 0, \quad 0 \leq p \leq P-1, \quad (31)$$

where

$$c^{(\mu,\mu')} = a_\mu / (a_\mu - b_{\mu'}),$$

$$a_\mu = e^{j2\pi z_\mu / Q},$$

$$b_{\mu'} = e^{j2\pi(i_{\mu'} + \Delta) / Q}.$$

It follows from (32) that  $\text{rank}\{\mathbf{\Lambda}^{(\mu,\mu')}\} = P$ . Let

$$\mathbf{\Lambda}_p = [\mathbf{I}_{N_t} \otimes (\mathbf{e}_P^p)^T] \mathbf{\Lambda} [\mathbf{I}_{N_t} \otimes (\mathbf{e}_P^p)^T]^T,$$

which denotes the  $N_t \times N_t$  sub-matrix of  $\mathbf{\Lambda}$ . Then, we can decompose  $\mathbf{\Lambda}_p$  as follows

$$\mathbf{\Lambda}_p = \frac{1}{\sqrt{Q}} (1 - e^{j2\pi\Delta}) \mathbf{\Lambda}_p^a \mathbf{C}_{N_t}^b \mathbf{\Lambda}_p^b, \quad (32)$$

where

$$\begin{aligned}\mathbf{A}_p^a &= \text{diag}\{a_0^{-(p)P/P}, \dots, a_{N_t-1}^{-(p)P/P}\}, \\ \mathbf{A}_p^b &= \text{diag}\{b_0^{-(p)P/P}, \dots, b_{N_t-1}^{-(p)P/P}\},\end{aligned}$$

and  $\mathbf{C}_{N_t}$  is the  $N_t \times N_t$  square matrix with its element given by  $[\mathbf{C}_{N_t}]_{\mu, \mu'} = c^{(\mu, \mu')}$ . From the definitions of  $a_\mu$  and  $b_{\mu'}$ , we have

$$\begin{aligned}a_\mu \neq 0 \ \& \ a_\mu \neq b_\mu \ \& \\ a_\mu \neq a_{\mu'}, \ b_\mu \neq b_{\mu'}, \ a_\mu \neq b_{\mu'}, \ \forall \mu \neq \mu'.\end{aligned}\quad (33)$$

Moreover, we also have (34) as shown at the bottom of the page. Then, with the assumption that  $N_t$  and  $Q$  are not very large (for example  $N_t \leq 8, Q = 16$ ), the determinant of  $\mathbf{C}_{N_t}$  can be obtained with its definition as follows

$$\begin{aligned}\det\{\mathbf{C}_{N_t}\} &= \prod_{\mu=1}^{N_t-1} \prod_{\mu'=0}^{\mu-1} \prod_{\mu''=0}^{N_t-1} \frac{(b_\mu - b_{\mu'})(a_{\mu'} - a_\mu)a_{\mu''}}{(a_\mu - b_{\mu'})(a_{\mu''} - b_\mu)(a_{\mu''} - b_{\mu'})} \\ &\neq 0,\end{aligned}\quad (35)$$

which shows that  $\text{rank}\{\mathbf{C}_{N_t}\} = N_t$ . From (32), we immediately obtain that  $\text{rank}\{\mathbf{A}_p\} = N_t$ . With the special diagonal structure of  $\mathbf{A}$ , we establish that  $\text{rank}\{\mathbf{A}\} = N_t P$  and then  $\text{rank}\{\mathbf{C}\} = N_t P$ . Exploiting the following relationship

$$\begin{aligned}\text{rank}\{\mathbf{A}_0\} + \text{rank}\{\mathbf{A}_1\} - n &\leq \text{rank}\{\mathbf{A}_0 \mathbf{A}_1\} \\ &\leq \min\{\text{rank}\{\mathbf{A}_0\}, \text{rank}\{\mathbf{A}_1\}\},\end{aligned}\quad (36)$$

where  $\mathbf{A}_0$  has  $n$  columns and  $\mathbf{A}_1$  has  $n$  rows, we further establish that  $\text{rank}\{\mathbf{C} \tilde{\mathbf{S}} \mathbf{F}\} = N_t L$  for any non-integer  $\Delta$  (recall that  $P \geq L$  in our design condition). Hence, for any  $\mathbf{h}^{(\nu, \mu)} (\neq \mathbf{0}_L)$ , we immediately obtain from its definition that  $\mathbf{d}_\nu(\Delta) \neq \mathbf{0}_{N-N_t P}$  and then  $\mathbf{d}(\Delta) \neq \mathbf{0}_{N_r(N-N_t P)}$ .

2) When  $\Delta$  is an integer: By exploiting the structure of  $\mathbf{C}$ ,  $\mathbf{d}_\nu(\Delta)$  can be transformed into the following equivalent form

$$\begin{aligned}\mathbf{d}_\nu(\Delta) &= [(\mathbf{d}^{(\nu, 0)}(\Delta))^T, (\mathbf{d}^{(\nu, 1)}(\Delta))^T, \\ &\dots, (\mathbf{d}^{(\nu, \omega)}(\Delta))^T, \dots, (\mathbf{d}^{(\nu, Q-N_t-1)}(\Delta))^T]^T,\end{aligned}\quad (37)$$

where

$$\begin{aligned}\mathbf{d}^{(\nu, \omega)}(\Delta) &= \\ &\sum_{\mu=0}^{N_t-1} \left\{ \mathbf{C}^{(\omega, \mu)} \text{diag}\{\tilde{\mathbf{s}}_\mu\} \Theta_{i_\mu}^T \mathbf{F}_N [\mathbf{I}_L, \mathbf{0}_{L \times (N-L)}]^T \mathbf{h}^{(\nu, \mu)} \right\}.\end{aligned}$$

For any integer  $\Delta \in (-Q, 0) \cup (0, Q)$ , with our design condition we have

$$(\mathbf{1}_Q - \mathbf{l})^T \mathbf{l}^{(\Delta)} > 0. \quad (38)$$

Without loss of generality, suppose  $(i_\mu + \Delta)_Q = z_\omega$ . Then,

together with condition (C0), which implies  $i_\mu = i_{\mu'}$  iff.  $\mu = \mu'$ , we have

$$\mathbf{C}^{(\omega, \mu)} = \sqrt{N} \mathbf{I}_P \ \& \ \mathbf{C}^{(\omega, \mu')} = \mathbf{0}_{P \times P}, \text{ if } \mu' \neq \mu. \quad (39)$$

Hence,

$$\begin{aligned}\mathbf{d}^{(\nu, \omega)}(\Delta) &= \\ &\sqrt{N} \text{diag}\{\tilde{\mathbf{s}}_\mu\} \Theta_{i_\mu}^T \mathbf{F}_N [\mathbf{I}_L, \mathbf{0}_{L \times (N-L)}]^T \mathbf{h}^{(\nu, \mu)}.\end{aligned}\quad (40)$$

With the condition  $P \geq L$ , we immediately obtain  $\mathbf{d}^{(\nu, \omega)}(\Delta) \neq \mathbf{0}_P$  and then  $\mathbf{d}(\Delta) \neq \mathbf{0}_{N_r(N-N_t P)}$  for any  $\mathbf{h}^{(\nu, \mu)} (\neq \mathbf{0}_L)$ .

Combining the above two cases, we draw the conclusion that  $\mathbf{d}(\Delta) \neq \mathbf{0}_{N_r(N-N_t P)}$  for any  $\mathbf{h}^{(\nu, \mu)} (\neq \mathbf{0}_L)$  and any  $\Delta \in (-Q, 0) \cup (0, Q)$ . This completes the proof.

## APPENDIX II

In this appendix, we will discuss the uniqueness of  $\hat{\varepsilon}_i = \varepsilon_i$  in maximizing the estimation metric for any  $\mathbf{h}^{(\nu, \mu)} (\neq \mathbf{0}_L)$ . For the case that  $\varepsilon = \varepsilon_i$ , it follows immediately from Theorem 1 that the uniqueness of  $\hat{\varepsilon}_i = \varepsilon_i$  in maximizing the estimation metric for any  $\mathbf{h}^{(\nu, \mu)} (\neq \mathbf{0}_L)$  is guaranteed. Therefore, in the following, we only consider the case that  $\varepsilon \neq \varepsilon_i$ , i.e.,  $\varepsilon_f \neq 0$ .

Define  $\Delta_i = \varepsilon - \varepsilon_i$ . Then, the estimation metric  $\mathfrak{P}(\Delta_i)$  can be written into an equivalent form as shown in (41) at the bottom of the next page. In (41),

$$\begin{aligned}\mathbf{M}^{(\mu, \mu', \mu'')}(\Delta_i) &= \mathbf{S}_\mu^H \mathbf{G}_{\mu'}(\Delta_i) \mathbf{S}_{\mu''}, \\ \mathbf{G}_\mu(\Delta_i) &= \mathbf{D}_N(-\Delta_i) \mathbf{F}_N^H \Theta_{i_\mu} \Theta_{i_\mu}^T \mathbf{F}_N \mathbf{D}_N(\Delta_i), \\ \mathbf{S}_\mu &= \sqrt{N} \mathbf{F}_N^H \Theta_{i_\mu} \text{diag}\{\tilde{\mathbf{s}}_\mu\} \Theta_{i_\mu}^T \mathbf{F}_N [\mathbf{I}_L, \mathbf{0}_{L \times (N-L)}]^T.\end{aligned}$$

From the definition of  $\mathbf{G}_\mu(\Delta_i)$ , we can obtain

$$\begin{aligned}\mathbf{G}_\mu(\Delta_i) &= \\ &\begin{bmatrix} \mathbf{G}_\mu^{(0, 0)}(\Delta_i) & \dots & \mathbf{G}_\mu^{(0, Q-1)}(\Delta_i) \\ \vdots & \mathbf{G}_\mu^{(q, q')}(\Delta_i) & \vdots \\ \mathbf{G}_\mu^{(Q-1, 0)}(\Delta_i) & \dots & \mathbf{G}_\mu^{(Q-1, Q-1)}(\Delta_i) \end{bmatrix},\end{aligned}\quad (42)$$

where

$$\mathbf{G}_\mu^{(q, q')}(\Delta_i) = Q^{-1} e^{j2\pi(q-q')(i_\mu - \Delta_i)/Q} \cdot \mathbf{I}_P.$$

Define

$$\mathbf{s}_\mu = \frac{1}{\sqrt{Q}} \mathbf{F}_P^H \tilde{\mathbf{s}}_\mu.$$

Then, we can also obtain that  $\mathbf{S}_\mu$  is an  $N \times L$  column-wise circulant matrix with  $[\mathbf{D}_Q(i_\mu P) \mathbf{1}_Q] \otimes [\mathbf{D}_P(i_\mu) \mathbf{s}_\mu]$  being its

$$\begin{aligned}\prod_{\mu=1}^{N_t-1} \prod_{\mu'=0}^{\mu-1} \{(b_\mu - b_{\mu'})(a_{\mu'} - a_\mu)\} &= \prod_{\mu=2}^{N_t-1} \prod_{\mu'=1}^{\mu-1} (b_\mu - b_{\mu'}) \\ &\times \sum_{\mu''=0}^{N_t-1} \left\{ (-1)^{\mu''} \cdot \prod_{\mu=1, \mu \neq \mu''}^{N_t-1} \prod_{\mu'=0, \mu' \neq \mu''}^{\mu-1} (a_{\mu'} - a_\mu) \cdot \prod_{\mu=1}^{N_t-1} (a_{\mu''} - b_\mu) \cdot \prod_{\mu=0, \mu \neq \mu''}^{N_t-1} (a_\mu - b_0) \right\}.\end{aligned}\quad (34)$$

first column vector. Hence, we have

$$[\mathbf{M}^{(\mu, \mu', \mu'')}(\Delta_i)]_{l, l'} = \alpha(\mu, \mu''; l, l')/Q \times \left\{ \sum_{q=0}^{Q-1} \sum_{q'=0}^{Q-q-1} \left\{ e^{jq'\psi} e^{jq\theta} + e^{j(q+q')\psi} e^{-jq\theta} \right\} - \sum_{q=0}^{Q-1} e^{jq\psi} \right\}, \quad (43)$$

where

$$\begin{aligned} \alpha(\mu, \mu''; l, l') &= e^{j2\pi(li_\mu - l'i_{\mu''})/N} \cdot (\mathbf{s}_\mu^{(l)})^H \mathbf{D}_P(i_{\mu''} - i_\mu) \mathbf{s}_{\mu''}^{(l')}, \\ \theta &= 2\pi(i_{\mu'} - i_\mu - \Delta_i)/Q, \\ \psi &= 2\pi(i_{\mu''} - i_\mu)/Q. \end{aligned}$$

When  $\mu = \mu''$ , imposing the condition that the elements of  $\tilde{\mathbf{s}}_\mu$  have constant amplitude, which translates to the time-domain zero auto-correlation (ZAC) property of  $\mathbf{s}_\mu$  [15], from (43) we immediately obtain (44) as shown at the bottom of the page. When  $\mu \neq \mu''$ , from (43) we obtain (45) as shown at the bottom of the page. Define

$$\zeta^{(\mu, \mu', \mu'')}(\Delta_i) = \left| [\mathbf{M}^{(\mu, \mu', \mu)}(\Delta_i)]_{l, l} / [\mathbf{M}^{(\mu, \mu', \mu'')}(\Delta_i)]_{l, l'} \right|.$$

Then, we immediately establish

$$\begin{aligned} \zeta^{(\mu, \mu', \mu'')}(\Delta_i) &= P/[N_t \cdot |\alpha(\mu, \mu''; l, l')|] \\ &\times |\sin(\psi/2)| \cdot |\cot(\psi/2) - \cot(\theta/2)|. \quad (46) \end{aligned}$$

The following remarks are now in order.

*Remark 1:* Both  $[\mathbf{M}^{(\mu, \mu', \mu)}(\Delta_i)]_{l, l}$  and  $[\mathbf{M}^{(\mu, \mu', \mu'')}(\Delta_i)]_{l, l'}$  are the period- $Q$  functions with respect to  $\Delta_i$ . Fix the true CFO  $\varepsilon$  and vary the candidate ICFO  $\tilde{\varepsilon}_i$  between  $(- \lfloor Q/2 \rfloor, Q - \lfloor Q/2 \rfloor]$ . Then, we have

$$\Delta_i \in [\varepsilon - Q + \lfloor Q/2 \rfloor, \varepsilon + \lfloor Q/2 \rfloor] \subset (-Q, Q). \quad (47)$$

For description convenience, we employ  $r(\Delta_i)$  to normalize  $\Delta_i$  from  $[\varepsilon - Q + \lfloor Q/2 \rfloor, \varepsilon + \lfloor Q/2 \rfloor]$  to  $[- \lfloor Q/2 \rfloor, Q - \lfloor Q/2 \rfloor]$ , where

$$\begin{aligned} r(\Delta_i) &= \Delta_i - [\text{sign}(\Delta_i - Q + \lfloor Q/2 \rfloor) \\ &\quad + \text{sign}(\Delta_i + \lfloor Q/2 \rfloor)]Q/2. \quad (48) \end{aligned}$$

Therefore,  $[\mathbf{M}^{(\mu, \mu', \mu)}(\Delta_i)]_{l, l}$  achieves its minimum 0 when  $r(\Delta_i) \in \{- \lfloor Q/2 \rfloor, - \lfloor Q/2 \rfloor + 1, \dots, Q - \lfloor Q/2 \rfloor - 1\} \setminus \{r(i_{\mu'} - i_\mu)\}$  and its maximum  $N/N_t$  when  $r(\Delta_i) = r(i_{\mu'} - i_\mu)$ . While  $|\mathbf{M}^{(\mu, \mu', \mu'')}(\Delta_i)]_{l, l'}|$  achieves its minimum 0 when  $r(\Delta_i) \in \{- \lfloor Q/2 \rfloor, - \lfloor Q/2 \rfloor + 1, \dots, Q - \lfloor Q/2 \rfloor - 1\}$ .

*Remark 2:* Both  $[\mathbf{M}^{(\mu, \mu', \mu)}(\Delta_i)]_{l, l}$  and  $[\mathbf{M}^{(\mu, \mu', \mu'')}(\Delta_i)]_{l, l'}$  are the continuous functions with respect to  $\Delta_i$ . For any  $\gamma > 0$  (for example  $\gamma = 10^{-2}$ ), there exists  $\delta > 0$  (for example  $\delta = 0.1$ ) which makes the relationships as shown in (49) and (50) at the bottom of the next page hold.

*Remark 3:* Assume  $0 < |\varepsilon_f| < 0.5$ . Then, for  $r(\varepsilon_i - \tilde{\varepsilon}_i) \in \{- \lfloor Q/2 \rfloor, - \lfloor Q/2 \rfloor + 1, \dots, Q - \lfloor Q/2 \rfloor - 1\}$ , we have (51) as shown at the bottom of the next page.

*Remark 4:*  $\zeta^{(\mu, \mu', \mu'')}(\Delta_i)$  is the period- $Q$  continuous function with respect to  $\Delta_i$ , which achieves its minimum 0 when  $r(\Delta_i) = r(i_{\mu'} - i_{\mu''})$  and its maximum  $+\infty$  when  $r(\Delta_i) = r(i_{\mu'} - i_\mu)$ . Impose the condition that  $|\alpha(\mu, \mu''; l, l')| < 1/N_t$  for  $\mu \neq \mu''$ . Then, there exists  $\chi > 0$  (for example  $\chi = 5$ ) which makes the relationship as shown in (52) at the bottom of the next page hold.

We use an example as shown in Fig. 5 to illustrate the above remarks.

Assume

$$\mathbb{E}\{[\mathbf{h}^{(\nu, \mu)}]_l^* [\mathbf{h}^{(\nu', \mu')}]_{l'}\} = 0, \quad \forall (\nu, \mu) \neq (\nu', \mu').$$

Then, it follows from Remark 2 and 4 that (53) and (54) as shown at the bottom of the page can be obtained, respectively. According to its definition,  $[\mathbf{M}^{(\mu, \mu', \mu)}(\Delta_i)]_{l, l}$  must be one of the  $Q$  possible values whatever  $i_\mu$  or  $i_{\mu'}$  is when we vary  $\tilde{\varepsilon}_i$  from  $(- \lfloor Q/2 \rfloor + 1)$  to  $(Q - \lfloor Q/2 \rfloor)$ . From Remark 3, we immediately obtain that

$$\begin{aligned} [\mathbf{M}^{(\mu, \mu', \mu)}(\varepsilon_f)]_{l, l} / [\mathbf{M}^{(\mu, \mu', \mu)}(\varepsilon_f + (i_{\mu'} - i_\mu)_Q)]_{l, l} &\gg 1, \\ \text{if } (i_{\mu'} - i_\mu)_Q &> 1, \quad (55) \end{aligned}$$

and the items involved in the right hand side of (54) achieve their maximum when  $r(\varepsilon_i - \tilde{\varepsilon}_i) = r(i_{\mu'} - i_\mu)$  for any  $\mathbf{h}^{(\nu, \mu)} (\neq \mathbf{0}_L)$ . Then, we can establish that there are  $N_r N_t$  items involved in the right hand side of (54) which achieve the maximum when  $\tilde{\varepsilon}_i = \varepsilon_i$  (i.e.,  $\mu = \mu'$ ), and at most  $N_r(N_t -$

$$\begin{aligned} \mathfrak{P}(\Delta_i) &= \sum_{\nu=0}^{N_r-1} \sum_{\mu=0}^{N_t-1} \sum_{\mu'=0}^{N_t-1} \left\{ (\mathbf{h}^{(\nu, \mu)})^H \mathbf{M}^{(\mu, \mu', \mu)}(\Delta_i) \mathbf{h}^{(\nu, \mu)} \right\} \\ &\quad + \sum_{\nu=0}^{N_r-1} \sum_{\mu=0}^{N_t-1} \sum_{\mu'=0}^{N_t-1} \sum_{\mu''=0, \mu'' \neq \mu}^{N_t-1} \left\{ (\mathbf{h}^{(\nu, \mu)})^H \mathbf{M}^{(\mu, \mu', \mu'')}(\Delta_i) \mathbf{h}^{(\nu, \mu'')} \right\}, \quad (41) \end{aligned}$$

$$\begin{aligned} \mathbf{M}^{(\mu, \mu', \mu)}(\Delta_i) &= \left\{ \frac{P}{N_t} + \frac{2P}{N_t Q} \Re \left\{ [(Q-1)e^{j\theta} - Qe^{j2\theta} + e^{j(Q+1)\theta}] / (1 - e^{j\theta})^2 \right\} \right\} \cdot \mathbf{I}_L \\ &= \frac{P}{N_t Q} \cdot \frac{1 - \cos(Q\theta)}{1 - \cos\theta} \cdot \mathbf{I}_L. \quad (44) \end{aligned}$$

$$\begin{aligned} [\mathbf{M}^{(\mu, \mu', \mu'')}(\Delta_i)]_{l, l'} &= \alpha(\mu, \mu''; l, l')/Q \cdot (1 - e^{jQ\theta})^2 / [(1 - e^{j\theta})(e^{j\psi} - e^{j\theta})e^{j(Q-1)\theta}] \\ &= \alpha(\mu, \mu''; l, l')/Q \cdot \frac{1 - \cos(Q\theta)}{\cos(\psi/2) - \cos(\theta - \psi/2)}. \quad (45) \end{aligned}$$

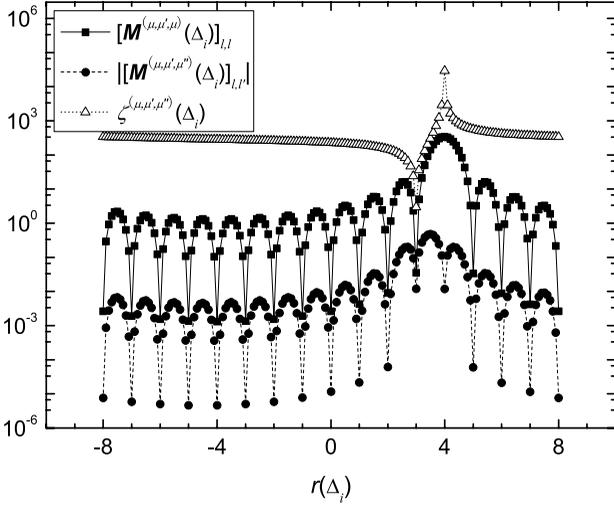


Fig. 5.  $[M^{(\mu,\mu',\mu)}(\Delta_i)]_{l,l}$ ,  $|[M^{(\mu,\mu',\mu'')}(\Delta_i)]_{l,l'}|$  and  $\zeta^{(\mu,\mu',\mu'')}(\Delta_i)$  versus  $r(\Delta_i)$  with  $P = 64$ ,  $Q = 16$ ,  $N_t = 3$ ,  $i_\mu = 3$ ,  $i_{\mu'} = 7$ ,  $i_{\mu''} = 4$ .

$\min_{1 \leq q \leq Q-1} \{(1_Q - \mathbf{l})^T \mathbf{l}^{(q)}\}$  items that achieve the maximum when  $\tilde{\varepsilon}_i \neq \varepsilon_i$  (i.e.,  $\mu \neq \mu'$ ). Due to the random snapshot channel energies, a closed form of training design conditions to yield the uniqueness of  $\hat{\varepsilon}_i = \varepsilon_i$  is intractable. However, it follows from condition (C3) and the above analysis that the reliability of the uniqueness can be maximized by designing the training sequences to maximize  $\min_{\mu' \neq \mu} \{(i_{\mu'} - i_\mu)_Q\}$  and

$$\min_{1 \leq q \leq Q-1} \{(1_Q - \mathbf{l})^T \mathbf{l}^{(q)}\}.$$

#### ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their valuable comments which helped to improve the quality of the paper.

#### REFERENCES

- [1] G. L. Stuber, J. R. Barry, S. McLaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.
- [2] P. Moose, "A technique for orthogonal frequency division multiplexing frequency offset correction," *IEEE Trans. Commun.*, vol. 42, pp. 2908–2914, Oct. 1994.
- [3] J. van de Beek, M. Sandell, and P. O. Borjesson, "ML estimation of time and frequency offset in OFDM systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 1800–1805, July 1997.

$$[M^{(\mu,\mu',\mu)}(\Delta_i)]_{l,l} < \gamma, \text{ if } r(\Delta_i) \in \bigcup_{q=-\lfloor Q/2 \rfloor}^{Q-\lfloor Q/2 \rfloor} \{(q-\delta, q) \cup (q, q+\delta)\} \setminus \{(r(i_{\mu'} - i_\mu) - \delta, r(i_{\mu'} - i_\mu) + \delta)\} \\ \setminus \{(-\lfloor Q/2 \rfloor - \delta, -\lfloor Q/2 \rfloor) \cup (Q - \lfloor Q/2 \rfloor, Q - \lfloor Q/2 \rfloor + \delta)\}, \quad (49)$$

$$|[M^{(\mu,\mu',\mu'')}(\Delta_i)]_{l,l'}| < \gamma, \text{ if } r(\Delta_i) \in \bigcup_{q=-\lfloor Q/2 \rfloor}^{Q-\lfloor Q/2 \rfloor} \{(q-\delta, q) \cup (q, q+\delta)\} \\ \setminus \{(-\lfloor Q/2 \rfloor - \delta, -\lfloor Q/2 \rfloor) \cup (Q - \lfloor Q/2 \rfloor, Q - \lfloor Q/2 \rfloor + \delta)\}. \quad (50)$$

$$[M^{(\mu,\mu',\mu)}(\Delta_i | r(\varepsilon_i - \tilde{\varepsilon}_i) = r(i_{\mu'} - i_\mu))]_{l,l} / [M^{(\mu,\mu',\mu)}(\Delta_i | r(\varepsilon_i - \tilde{\varepsilon}_i) \neq r(i_{\mu'} - i_\mu))]_{l,l} \\ = \sin^2 \{ \pi [\varepsilon_f + r(\varepsilon_i - \tilde{\varepsilon}_i) - r(i_{\mu'} - i_\mu)] / Q \} / \sin^2 (\pi \varepsilon_f / Q) \Big|_{r(\varepsilon_i - \tilde{\varepsilon}_i) \neq r(i_{\mu'} - i_\mu)} > 1. \quad (51)$$

$$\zeta^{(\mu,\mu',\mu'')}(\Delta_i) > P \cdot |\sin(\psi/2)| \cdot |\cot(\psi/2) - \cot(\psi/2 + \pi\delta/Q)| > \chi, \\ \text{if } r(\Delta_i) \in \{(-\lfloor Q/2 \rfloor, r(i_{\mu'} - i_{\mu''}) - \delta) \cup (r(i_{\mu'} - i_{\mu''}) + \delta, Q - \lfloor Q/2 \rfloor)\}. \quad (52)$$

$$\mathfrak{P}(\Delta_i) \doteq 0, \text{ if } r(\Delta_i) \in \bigcup_{q=-\lfloor Q/2 \rfloor}^{Q-\lfloor Q/2 \rfloor} \{(q-\delta, q) \cup (q, q+\delta)\} \setminus \left\{ \bigcup_{\mu,\mu'=0}^{N_t-1} \{(r(i_{\mu'} - i_\mu) - \delta, r(i_{\mu'} - i_\mu) + \delta)\} \right\} \\ \setminus \{(-\lfloor Q/2 \rfloor - \delta, -\lfloor Q/2 \rfloor) \cup (Q - \lfloor Q/2 \rfloor, Q - \lfloor Q/2 \rfloor + \delta)\}, \quad (53)$$

$$\mathfrak{P}(\Delta_i) \doteq \sum_{\nu=0}^{N_r-1} \sum_{\mu=0}^{N_t-1} \sum_{\mu'=0}^{N_t-1} \left\{ (\mathbf{h}^{(\nu,\mu)})^H \mathbf{M}^{(\mu,\mu',\mu)}(\Delta_i) \mathbf{h}^{(\nu,\mu)} \right\}, \\ \text{if } r(\Delta_i) \in \bigcup_{q=-\lfloor Q/2 \rfloor}^{Q-\lfloor Q/2 \rfloor-1} \{[q+\delta, q+1-\delta]\} \cup \left\{ \bigcup_{\mu,\mu'=0}^{N_t-1} \{(r(i_{\mu'} - i_\mu) - \delta, r(i_{\mu'} - i_\mu) + \delta)\} \right\}. \quad (54)$$

- [4] U. Tureli, H. Liu, and M. D. Zoltowski, "OFDM blind carrier offset estimation: ESPRIT," *IEEE Trans. Commun.*, vol. 48, pp. 1459–1461, Sept. 2000.
- [5] M. Morelli and U. Mengali, "An improved frequency offset estimator for OFDM applications," *IEEE Commun. Lett.*, vol. 3, pp. 75–77, Mar. 1999.
- [6] D. Huang and K. B. Letaief, "Carrier frequency offset estimation for OFDM systems using null subcarriers," *IEEE Trans. Commun.*, vol. 54, no. 5, pp. 813–823, May 2006.
- [7] —, "Enhanced carrier frequency offset estimation for OFDM using channel side information," *IEEE Trans. Wireless Commun.*, vol. 5, no. 10, pp. 2784–2793, Oct. 2006.
- [8] Y. Yao and G. B. Giannakis, "Blind carrier frequency offset estimation in SISO, MIMO and multiuser OFDM systems," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 173–183, Jan. 2005.
- [9] A. N. Mody and G. L. Stuber, "Synchronization for MIMO OFDM systems," in *Proc. IEEE Globecom'01*, vol. 1, Nov. 2001, pp. 509–513.
- [10] A. van Zelst and T. C. W. Schenk, "Implementation of a MIMO OFDM based wireless LAN system," *IEEE Trans. Signal Processing*, vol. 52, pp. 483–494, Feb. 2004.
- [11] O. Besson and P. Stoica, "On parameter estimation of MIMO flat-fading channels with frequency offsets," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 602–613, Mar. 2003.
- [12] X. Ma, M. K. Oh, G. B. Giannakis, and D. J. Park, "Hopping pilots for estimation of frequency-offset and multi-antenna channels in MIMO OFDM," *IEEE Trans. Commun.*, vol. 53, no. 1, pp. 162–172, Jan. 2005.
- [13] F. Simoens and M. Moeneclaey, "Reduced complexity data-aided and code-aided frequency offset estimation for flat-fading MIMO channels," *IEEE Trans. Wireless Commun.*, vol. 5, no. 6, pp. 1558–1567, June 2006.
- [14] P. Stoica and O. Besson, "Training sequence design for frequency offset and frequency-selective channel estimation," *IEEE Trans. Commun.*, vol. 51, no. 11, pp. 1910–1917, Nov. 2003.
- [15] H. Minn, X. Fu, and V. K. Bhargava, "Optimal periodic training signal for frequency offset estimation in frequency-selective fading channels," *IEEE Trans. Commun.*, vol. 54, no. 6, pp. 1081–1096, June 2006.
- [16] M. Ghogho and A. Swami, "Training design for multipath channel and frequency offset estimation in MIMO systems," *IEEE Trans. Signal Processing*, vol. 54, no. 10, pp. 3957–3965, Oct. 2006.
- [17] H. Minn, N. Al-Dhahir, and Y. Li, "Optimal training signals for MIMO OFDM channel estimation in the presence of frequency offset and phase noise," *IEEE Trans. Commun.*, vol. 54, no. 10, pp. 1754–1759, Oct. 2006.
- [18] H. Minn and N. Al-Dhahir, "Optimal training signals for MIMO OFDM channel estimation," *IEEE Trans. Wireless Commun.*, vol. 5, no. 5, pp. 1158–1168, May 2006.
- [19] D. Chu, "Polyphase codes with good periodic correlation properties," *IEEE Trans. Inform. Theory*, vol. 18, pp. 531–532, July 1972.
- [20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.
- [21] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 5, pp. 720–741, May 1989.
- [22] M. Pesavento, A. B. Gershman, and M. Haardt, "Unitary root-MUSIC with a real-valued eigendecomposition: a theoretical and experimental performance study," *IEEE Trans. Signal Processing*, vol. 48, no. 5, pp. 1306–1314, May 2000.
- [23] Y. X. Jiang, X. Q. Gao, X. H. You, and W. Heng, "Training sequence assisted frequency offset estimation for MIMO OFDM," in *Proc. IEEE ICC'06*, vol. 12, June 2006, pp. 5371–5376.
- [24] W. H. Press, *Numerical Recipes in C++: the Art of Scientific Computing*. Cambridge University Press, 2002.
- [25] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The John Hopkins University Press, 1996.
- [26] F. Gao and A. Nallanathan, "Blind maximum likelihood CFO estimation for OFDM systems via polynomial rooting," *IEEE Signal Processing Lett.*, vol. 13, no. 2, pp. 73–76, Feb. 2006.
- [27] W. H. Mow, "A new unified construction of perfect root-of-unity sequences," in *Proc. IEEE Spread-Spectrum Techniques and Applications*, vol. 3, Sept. 1996, pp. 955–959.
- [28] J. Coon, M. Beach, and J. McGeehan, "Optimal training sequences for channel estimation in cyclic-prefix-based single-carrier systems with transmit diversity," *IEEE Signal Processing Lett.*, vol. 11, no. 9, pp. 729–732, Sept. 2004.
- [29] F. Gini and R. Reggiannini, "On the use of Cramer-Rao-like bounds in the presence of random nuisance parameters," *IEEE Trans. Commun.*, vol. 48, no. 12, pp. 2120–2126, Dec. 2000.

**Yanxiang Jiang** (S'03-M'08) received the B.S. degree in electrical engineering from Nanjing University, Nanjing, China, in 1999, the M.E. degree and the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 2003 and 2007, respectively.

He received the NJU excellent graduate honor from Nanjing University in 1999. His research interests include mobile communications, wireless signal processing, parameter estimation, synchronization, signal design, and digital implementation of communication systems.

**Hlaing Minn** (S'99-M'01-SM'07) received his B.E. degree in Electronics from Yangon Institute of Technology, Yangon, Myanmar, in 1995, M.Eng. degree in Telecommunications from Asian Institute of Technology (AIT), Pathumthani, Thailand, in 1997 and Ph.D. degree in Electrical Engineering from the University of Victoria, Victoria, BC, Canada, in 2001.

He was with the Telecommunications Program in AIT as a laboratory supervisor during 1998. He was a research assistant from 1999 to 2001 and a post-doctoral research fellow during 2002 in the Department of Electrical and Computer Engineering at the University of Victoria. Since September 2002, he has been with the Erik Jonsson School of Engineering and Computer Science, the University of Texas at Dallas, USA, as an Assistant Professor. His research interests include wireless communications, statistical signal processing, error control, detection, estimation, synchronization, signal design, and cross-layer design. He is an Editor for the *IEEE Transactions on Communications*.

**Xiqi Gao** (SM'07) received the Ph.D. degree in electrical engineering from Southeast University, Nanjing, China, in 1997. He joined the Department of Radio Engineering, Southeast University, in April 1992. Now he is a professor of information systems and communications. From September 1999 to August 2000, he was a visiting scholar at Massachusetts Institute of Technology, Cambridge, and Boston University, Boston, MA. His current research interests include broadband multi-carrier transmission for beyond 3G mobile communications, space-time wireless communications, iterative detection/decoding, signal processing for wireless communications.

Dr. Gao received the Science and Technology Progress Awards of the State Education Ministry of China in 1998 and 2006. He is currently serving as an editor for the *IEEE Transactions on Wireless Communications*.

**Xiaohu You** (M'91) received the M.S. and Ph.D. degrees in electrical engineering from Southeast University, Nanjing, China, in 1985 and 1988, respectively.

Since 1990 he has been working with National Mobile Communications Research Laboratory at Southeast University, where he holds the ranks of professor and director. From 1993 to 1997 he was engaged, as a team leader, in the development of China's first GSM and CDMA trial systems. He was the Premier Foundation Investigator of the China National Science Foundation in 1998. From 1999 to 2001 he was on leave from Southeast University, working as the chief director of China's 3G (C3G) Mobile Communications R&D Project. He is currently responsible for organizing China's B3G R&D activities under the umbrella of the National 863 High-Tech Program, and he is also the chairman of the China 863-FUTURE Expert Committee. He has published two books and over 20 IEEE journal papers in related areas. His research interests include mobile communications, advanced signal processing, and applications.

**Yinghui Li** (S'05) received the B.E. and M.S. degree in Electrical Engineering from the Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2000 and 2003, respectively, and the Ph.D. degree in Electrical Engineering from the University of Texas at Dallas in 2007. Her research interests are in the applications of statistical signal processing in synchronization, channel estimation and detection problems in broadband wireless communications.